



territoires, environne-  
ment, télédétection  
et information spatiale

vincent douzal

Philippe Lemoisson

+33 04 67 54 87 59

+33 04 67 59 37 42

vincent.douzal@teledetection.fr

philippe.lemoisson@cirad.fr

références à rappeler :

vos références :

pièces jointes :

imprimé le :

Montpellier,

027.6 SddeC

schéma directeur pour les données de l'environnement de la Corse

mercredi 2 septembre 2009 14:35:00

le vendredi 24 avril 2009

M. Jacques Nicolau  
Dreal Corse

copies à

**note de synthèse : schéma directeur pour les données de l'environnement de la Corse**

Il nous paraît important de situer le schéma en place en Corse, parmi l'ensemble des dispositifs contemporains.

### **–1 Une solution très avancée, et opérationnelle dans un paysage complexe, fortement contraint par les standards**

Les exigences d'une gestion environnementale durable, et leur traduction en de nouvelles dispositions réglementaires, conduisent les acteurs de la gestion des territoires à organiser un large accès du grand public aux données d'environnement. C'est dans cette perspective qu'une démarche et un dispositif ont été mis en place de façon itérative par la Diren, devenue Dreal de Corse, en élargissant progressivement le cercle d'acteurs, fournisseurs et consommateurs de données en partant de ses partenaires les plus proches. À l'obligation réglementaire de diffuser publiquement l'information, répond le besoin des services d'en disposer opérationnellement : la nécessité de mutualiser les données et réunir les moyens des multiples services en un ensemble fonctionnel ne se discute plus. Les deux peuvent et doivent être satisfaites par un seul et même dispositif.

Une organisation qui ne sera pas un nœud vivant de circulation de données, et notamment de celles qu'elle produit, et qui fera le choix de thésau-

riser, sera marginalisée. Incidemment, elle sera aussi au ban des exigences réglementaires.

Le dispositif mis en place à Ajaccio est à considérer dans un paysage riche et diversifié de systèmes, portails et logiciels, tous appelés à converger vers des formats et des protocoles d'échange internationalement standardisés. Les processus d'évolution concomitants, lents et délicats, exigent d'importants efforts de concertation, comparaison, confrontation, mise en phase.

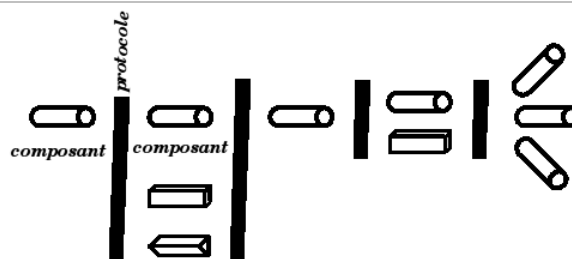


Figure 1 — Tous les systèmes homologues à celui mis en place en Corse visent à se conformer à un même jeu de standards. Ces systèmes sont alors scindés en une série de **composants** fonctionnellement indépendants, qui néanmoins fonctionnent de façon cohérente en se procurant les uns les autres des **services** mutuels, en communiquant selon des **protocoles** strictement, internationalement standardisés. Grâce à l'universalité de l'accord sur ces protocoles, des composants d'origines diverses communiquent entre eux et collaborent fluidement, ils ignorent même l'identité des composants avec lesquels ils échangent des services, ils savent seulement qu'il se conforment au même langage d'échange. Les composants qui partagent un même protocole sont dits **interopérables**. Le composant demandeur est classiquement baptisé **client**, celui qui lui répond est un **serveur**. L'architecture générale résultante est dite en **couches** (tiers en anglais, c'est-à-dire étages), bien qu'elle puisse se ramifier, et plus encore que ne le suggère la figure

Cette architecture plébiscite une morale établie par une cinquantaine d'années de travaux et d'essais en informatique, à savoir que dans un système, « **les données dominent** » : leur organisation, leur structuration détermine ce qu'il est possible d'obtenir du système. Toute cette configuration en couches est effectivement commandée par les structures de données échangées, **présentées** dans des langages d'échange, les protocoles.

Le mouvement de standardisation est parti des protocoles les plus fondamentaux d'échanges et de routage sur les réseaux, et en vient maintenant à s'intéresser au détail **sémantique** des données que s'échangent des services hautement spécialisés dans un **métier** donné. On note par exemple les efforts de normalisation du vocabulaire sur l'eau conduits en France par l'Onema (Office national de l'eau et des milieux aquatiques) et mis en œuvre dans le Sandre (service d'administration nationale des données et référentiels sur l'eau). Ce processus, lent et délicat, puisqu'il peut imposer le démontage d'applications à architecture monolithique, exige d'importants efforts de concertation, comparaison, confrontation, mise en phase.

Certains systèmes sont relativement peu avancés dans ce mouvement migratoire. Par exemple, Cartorisque présente au public un panorama général des données et cartes sur les risques en France ; cependant à ce jour, l'alimentation des données est asynchrone, les producteurs de données transmettent épisodiquement les lots et la consultation passe souvent par un téléchargement préalable qui rompt la **chaîne des services**. Cette situation en fait essentiellement un stock de données mortes, une sorte de cul-de-sac : superficiellement, on obtient des indications, mais dès qu'on a un besoin précis, il excède ce qu'on peut obtenir des données disponibles.

C'est parce qu'un ensemble de protocoles instaurant un schéma de découpe en composants qui assurent des services déterminés est en place aujourd'hui et continue à se développer, que toute tentative de solution qui ne respecterait pas ces normes d'interopérabilité serait aujourd'hui vouée à l'échec. Grâce à cette standardisation, un ministère peut investir dans le développement de logiciels qui réalisent un tel ensemble de composants sans être à la merci d'une offre commerciale exclusive — c'est ce qu'a fait

celui en charge de l'écologie. Il a la certitude de la pérennité de son investissement, de ne jamais voir ses solutions marginalisées, ni son système totalement remis en question. Qui plus est, le fait qu'il rende le code de ses logiciels libre contribue à asseoir la vigueur des standards adoptés et garantit aussi la durabilité de l'investissement.

Cependant, à ce jour, on est encore dans **un vaste mouvement de migration** simultanée des systèmes d'information et des offres logicielles. D'une part la multitude de systèmes d'information, que tous les types d'organisations publiques ou privées entretiennent, évolue vers ces architectures de services. D'autre part les offres logicielles elles-mêmes migrent vers une conformation complète à l'immense palette de protocoles qui sont établis ou en cours de définition. Ces mutations simultanées rendent certaines analyses délicates.

Conformément aux considérations de la figure 1, le dispositif Corse doit s'analyser en **deux composantes** : d'une part, **des données**, présentées sous forme de services standard (WMS, WFS, bases SQL, etc.), ou, au minimum, de formats standards, dont la description est publique. D'autre part, **une solution logicielle**, comprenant un outil de cartographie dynamique et un catalogue, qui permet de suppléer au manque d'équipement d'un utilisateur (un « client » des services) en logiciel de système d'information géographique (SIG) par exemple. Il lui suffit alors d'un ordinateur muni d'un navigateur web pour avoir un plein accès aux données, y compris cartographiques, ce qui, si on lui en confère le droit, inclut la capacité à les modifier.

La solution logicielle en place en Corse est dans la conformité la plus complète qui soit à ce jour avec les standards concernant les échanges de données, en particulier géographiques. Relativement à d'autres offres<sup>1</sup>, elle possède quelques avances techniques, par exemple la capacité à procurer un véritable SIG en ligne, avec l'accès aux bases de données sous-jacentes, aux entités géographiques, mais avec aussi la possibilité de les mettre à jour, ce qui permet de répartir les personnes qui font ce travail souplement sur un territoire quelconque, pourvu que des moyens de communication existent.

La solution logicielle qui sous-tend le dispositif mis en place à Ajaccio est un produit du commerce, ArcGIS d'Esri, en appui sur une base SQL server. Cette solution étant conforme aux standards établis, elle permettrait d'envisager une mutation vers des solutions libres, ou autres<sup>2</sup>. Cependant, une telle mutation semble ne présenter aucun intérêt à un horizon visible à ce jour. D'une part, l'investissement en licences est fait, et les charges de maintenance locale ne seraient pas supprimées pour autant. D'autre part, la solution est aujourd'hui pleinement opérationnelle, et les services attendus par les utilisateurs finals sont rendus, ce qui est la raison d'être du dispositif<sup>3</sup>. Migrer vers un autre dispositif logiciel serait un travail important, qui immobiliserait l'équipe et tous les partenaires pendant une période sensible, pour aboutir au mieux au même niveau de fonctionnalité final (et, comme on l'a noté plus haut, probablement à un niveau de fonctionnalité inférieur).

---

<sup>1</sup> Par exemple l'offre Carmen, supportée par le ministère en charge de l'écologie, comparaison faite à la date de la rédaction.

<sup>2</sup> Une telle migration serait d'autant plus facilement réalisable que les idiosyncrasies qu'offrent ces logiciels, qui sont autant de pièges non-standards, ont été attentivement et soigneusement évitées par le concepteur lors de l'implémentation. Ainsi, le dictionnaire des données de l'infrastructure Corse est-il enregistré dans une table standard. Les créateurs de logiciels commerciaux ne manquent jamais, tout en se tenant aux standards établis par le marché, d'offrir des « améliorations » par lesquelles ils se donnent des chances de garder leurs clients captifs. De même les « géodatabases », de format privé ESRI, ne sont utilisées que comme des contenants pratiques pour des lots de données de référence.

<sup>3</sup> Ils sont même rendus, ou tout au moins offerts, bien au-delà de ce qui est effectivement entré en usage au près des utilisateurs, qui pour y venir pleinement doivent ajuster leur propre système pour qu'il se prête mieux à une architecture ouverte, en services. Ces ajustements, même mineurs, interagissent avec les politiques propres de chaque service qui ne sont pas encore conduites en harmonie, spécialement s'agissant des services déconcentrés de ministères divers, d'où un gros travail de mise en phase et discussion.

Or, comme on y reviendra au§-4, le cœur du travail se situe ailleurs, dans les données elles-mêmes.

## –2 Une grille de cinq critères pour évaluer des systèmes apparentés au schéma Corse

Nous proposons de dégager cinq critères permettant de fonder une éventuelle comparaison entre des « dispositifs » dédiés au partage d'informations dans un contexte où les sources comme les usages sont multiples. Ces critères se déploient dans des dimensions largement indépendantes, qui néanmoins se recoupent et se connectent ; l'ordre dans lequel ils sont présentés ci-dessous n'est pas significatif :

1. **consensus autour de la signification des données** : les données ne deviennent des informations que dans la mesure où elles font sens pour des acteurs. La question délicate du partage du sens se met en place au cours d'un processus qui fait partie de l'activité d'**administration des données**. Évaluer un dispositif sous cet angle passe par la réponse aux questions suivantes :

- y a-t-il un catalogue de métadonnées ?
- y a-t-il un dictionnaire des données stockées ?
- y a-t-il un vrai index ?
- leur processus de construction a-t-il bien impliqué tous les acteurs, et
- y a-t-on bien intégré et décrit leurs **schémas d'usage** des données, qui ultimement seuls en certifient la signification (sinon, le sens des données est enfoui dans le formalisme des descriptions de structures de données, qui le rendent hermétique.) ?

2. **propriété, contrôle, responsabilité vis-à-vis des données** : il est communément admis aujourd'hui que la donnée doit rester « au plus près de son producteur » ; ce qui signifie qu'il a tout contrôle sur leur publication et leur diffusion. C'est une traduction du principe de **subsidiarité**. L'idée de proximité doit être traduite dans un sens fonctionnel : le disque où sont enregistrées les données n'a pas à être à proximité de l'opérateur, mais celui-ci doit avoir toute latitude pour les manipuler, et la maîtrise complète de leur service de mise à disposition.

Le **droit** qui rend désormais obligatoire l'instauration de ces dispositifs pour rendre les données publiques vient se superposer à **d'autres régimes juridiques** qui lui placent des limites, par exemple, le code civil qui pose le respect de la vie privée. On peut s'interroger sur les responsabilités qui peuvent découler de la publication de données. Mais la diversité des acteurs et des situations dans lesquelles peut survenir un dommage qui se raccorde au droit de l'environnement, et peut donc par exemple mettre en cause des données de l'environnement est telle, avec une telle imbrication potentielle de régimes juridiques, qu'il est douteux de présumer en prévoir l'étendue exacte.

Par ailleurs, on se trouve dans une situation de **régime juridique non établi**, une situation inédite, dont le droit propre n'est pas constitué, puisqu'il croise deux domaines eux-mêmes nouveaux, à savoir la transparence des données de l'environnement et la disponibilité de vastes jeux de données, avec des facilités de manipulation inédites, grâce à leur forme électronique. La possibilité d'intégrer, de combiner des données (la vitesse, la facilité avec laquelle on le fait) fait exploser la combinatoire des assemblages, des recoupements qui peuvent être produits chaque jour, combinatoire elle-même amplifiée par de nouvelles modalités de croisements. L'utilisateur naïf peut se croire fondé à opérer tous les assemblages techniquement possibles. Mais de même que toutes les phrases grammaticalement correctes de la langue française ne sont pas sensées, toutes ces combinaisons n'ont pas nécessairement légitimes. On ne peut ni définir par avance ce qu'il est valide de faire, ni énumérer tout ce qui est exclu. Comme dans toute situation similaire, face aux inquiétudes que peut susciter ce vide, il n'existe pas de meilleure protection que de faire platement état de la façon dont sont obtenues les données, et l'**usage** auquel elles sont desti-

nées, qui est celui pour lequel elles ont été testées, donc validées. C'est d'ailleurs ce que demanderait un juge qui aurait à régler un litige.

Des questions essentielles sont donc :

- les questions de droit sont-elles abordées dans le processus de mutualisation ?
- Est-ce que ces questions sont proprement arrimées aux processus d'acquisition des données, à l'**usage** qui en est fait, en documentant soigneusement ces deux facettes ?
- Est-ce qu'un avertissement liminaire, et un texte adéquat plus développé permet aux utilisateurs de prendre la mesure des questions soulevées par la manipulation de données, en l'incitant à se plonger dans la documentation des données ?

**3. qualité des données** : cette notion n'est pas toujours facile à appréhender, la non-qualité pouvant provenir non seulement d'erreurs de mesure ou de saisie mais également (et peut-être surtout) d'une ambiguïté relative à l'**usage** qui peut être fait de l'information (point 1.). La meilleure méthode pour se prémunir de l'erreur est la « boucle de rétroaction » selon laquelle le producteur de la donnée est soit son premier utilisateur, soit en relation directe avec ses utilisateurs. La meilleure méthode pour se prémunir de l'ambiguïté est la description précise de la façon dont la donnée a été mesurée (la chaîne d'acquisition des données). C'est la meilleure qualification, la meilleure documentation que l'on puisse apporter en vue de sceller le point 1. Ainsi à la qualité des données, substitue-t-on une **qualification des données**. Les données, donc, sont-elles pleinement qualifiées ?.

**4. accessibilité des données** : il ne suffit pas que des données de qualité, dont la signification fait l'objet d'un consensus, existent dans un dispositif ; encore faut-il que tous ceux qui en sont en droit d'en avoir l'usage puissent y accéder et les manipuler sans devoir souscrire à une configuration particulière (autre que de se conformer aux standards en vigueur). Un dispositif fournit des données techniquement accessibles quand il obéit aux standards, normes techniques d'interopérabilité.

**5. pérennité des données** : cette dimension est trop souvent laissée pour compte et pourtant son importance est fondamentale, particulièrement lorsque les données gérées doivent permettre d'observer et décrire des évolutions, de mesurer des tendances à long terme, de servir de référence en cas de litige ultérieur. Il faut assurer la pérennité de l'ensemble de la chaîne technico-logicielle qui permet de lire les données, de façon à respecter, maintenant que les données sont exclusivement sous forme numérique, les principes présidant à la loi sur les archives classiques. Évaluer un dispositif sous cet angle passe par la réponse aux questions suivantes :

- y a-t-il uniquement des formats standards de stockage basé sur des normes ouvertes ?
- y a-t-il sécurisation des serveurs, réplication des données, sur des sites distincts et distants, sauvegardes régulières ?
- l'organisation des bases, la structure du système permet-elle de conserver et d'étudier des chroniques, y compris à long terme (ce qui protège aussi de pertes majeures) ?

Cette grille est un support de lecture, pour quiconque déchiffre un système d'information. Elle est présente en permanence à l'esprit de quelqu'un qui s'implique dans la conception, l'animation, la maintenance d'un système, spécialement un système mutualisé.

— **Il apparaît que la question de l'usage des données est le fait organisateur du système.** Largement développée dans le rapport 025, SddeC, elle est soulevée, sous une forme ou une autre, sur les cinq points d'analyse, et, quand on observe un système, sur ses aspects les plus abstraits comme ceux les plus pratiques. Par exemple, la prise en compte des conditions d'usage effectif, du travail concret effectué par les opérateurs qui nourrissent le jeu de données est essentielle pour assurer un système efficace, qui incite à maintenir des données de qualité, à jour. L'absence d'usage, ou la déconnexion entre l'acquisition et l'usage permet

d'identifier des pans de données sans qualité, sans qualification, on parle couramment de cimetières de données.

— Concrètement, **l'avertissement sur l'usage des données** figurant sur l'atlas des données de l'environnement de Corse est, à ce stade, une information adéquate ; seule la mention stipulant que les données sont fournies à titre informatif, sans valeur réglementaire doit être amendée, puisqu'elles ont une certaine valeur. Cependant, il faut préciser qu'elles ne sauraient se substituer aux procédures d'information ou d'autorisation, comme celles du fait de la loi sur l'eau, ou les permis de construire. **La nouveauté de cette situation de droit incite à en faire l'objet d'une étude spécifique.**

### **—3 De cette grille à une comparaison approfondie des systèmes**

On pourrait être tenté, sur la base de cette grille d'analyse, de comparer entre eux différents dispositifs, et notamment un certain nombre de ceux qui existent, survivent ou se développent en France (Carmen, Prodige, le Géo-web et le Géorépertoire du ministère de l'agriculture, Cartorisque, Cartélie, etc.). Mais conduire une telle analyse supposerait en premier lieu un accès à ces dispositifs pour pouvoir les appréhender de l'intérieur, et en second lieu de pratiquer de façon approfondie les manipulations courantes d'alimentation en données, d'administration, d'usage. Il s'agit évidemment d'un travail conséquent, qui serait par ailleurs extrêmement péremptible car tous ces systèmes sans exception, s'ils ne sont pas en cours d'abandon, sont en pleine migration vers une structure plus achevée en services telle qu'évoquée plus haut.

Accéder à l'intérieur de ces systèmes semble aller de soi, s'ils répondent aux principes généraux qui sont réalisés en Corse. Mais les niveaux de migration sont très différents. Prodige, par exemple, n'est absolument pas évaluable du point de vue d'un citoyen utilisateur, car il n'est accessible, à ce jour, même en simple consultation, qu'aux services de l'État, et parmi eux, à un sous-ensemble restreint de personnes accréditées. C'est là que la bât blesse : clos, il ne répond pas aux exigences des textes réglementaires sur l'accès à l'information cités dans le rapport 025.SddeC, il est du même coup hermétique à une exploration et a fortiori une évaluation publique.

L'impossibilité de cette comparaison détaillée n'empêche pas de se former une idée opérante sur un système. Il faut adopter un mode de raisonnement où l'on admet que l'on ne peut contrôler complètement un processus, ni la direction qui sera retenue à l'avenir, et où l'on contrôle que l'on se trouve dans le bon mouvement. Par ailleurs, il est impossible d'avancer seul, tout comme de laisser un partenaire sur le bord du chemin.

Enfin, il faut se convaincre que quelle que soit la sophistication des technologies et des méthodes employées, rien du système d'information n'existe de façon tangible, en l'absence de données. Cette domination des données achève de dissuader de tenter une comparaison entièrement développée. La comparaison ne pourrait s'arrêter à la surface des principes énoncés. Il y a obligation de plonger dans les modèles conceptuels de données, puis dans le contenu même de ces données, d'examiner à l'aune de nos critères chacune de leurs sources, chacun de leurs usages effectifs qui conditionnent toute la signification qu'elles revêtent. Au final, il ne s'agit pas d'un point de passage, d'une étape à franchir ; c'est **la matière même du schéma**, l'activité quotidienne de l'administration de ses données :

### **—4 Les données, leur inventaire, leur documentation, concentrent l'essentiel du travail**

C'est d'abord une conséquence de ce que les **données dominent**. C'est aussi parce que les données touchent à la matière même des métiers (point 1) et à la façon dont ils s'exercent. Il y a un vaste travail de refonte, de reconception qui accompagne le crible des données. Il faut aller au cœur des données, des usages, des besoins. Et comme les questions que l'on se

pose sur l'environnement sont en renouvellement perpétuel, ce travail sur les données se mutera en une activité entretenue, permanente, de même qu'aucun logiciel ne peut perdurer sans maintenance. Le rapport 025. SddeC parle d'inventaire permanent, à outiller et animer.

Le travail sur les données, qui consiste à s'entendre sur leur sémantique, c'est-à-dire établir des protocoles sur leur sémantique, et à documenter ce processus, est donc à la fois l'urgence, la priorité, et le gros du travail. C'est aussi l'activité essentielle du schéma.

Un tel travail d'inventaire et d'analyse des lots de données est au cœur du processus en cours dans le schéma actuellement en vigueur à Ajaccio. L'évolution continue des besoins, des usages sur les données implique par ailleurs que ce travail devienne un inventaire permanent.

Les partenaires qui ont un besoin mutuel doivent instituer un organe qui concentre et confronte les besoins et les avis de tous, et établit en commun un protocole (c'est-à-dire, quel que soit le sujet, une entente sur un vocabulaire, et la façon de former des phrases valides). Cette nécessité d'unification relève du **principe de suppléance**, l'exact miroir du principe de subsidiarité qui permet, l'entente acquise, à chacun de s'appuyer dessus pour assurer sa part de fonctionnement. C'est un processus spontané, et non pas une autorité classique, qui peut élire un partenaire comme moteur de cette suppléance, selon un processus. tout à fait similaire au fonctionnement du monde du logiciel libre.

Dans le périmètre français, et sur l'eau, on a vu que l'Onema a été instauré, à l'appel d'un ensemble d'acteurs, à partir d'un certain moment, pour assurer ce rôle de suppléance. La suppléance n'épargne à aucun des acteurs un **travail d'implication**, au minimum participer aux travaux communs. Elle ne peut non plus leur économiser l'**effort de spécialisation** ; qui consiste à produire, à proposer les éléments qui correspondent à une compétence qu'il est le seul à détenir, des missions qu'il est le seul à exercer, un usage qui lui est propre, une fois que cette spécificité a été bien délimitée collectivement. Cette délimitation consiste à scinder le domaine de travail en protocoles, articulant des composants (figure 1).

Pour ce qui est des données de l'environnement en Corse, la Dreal joue naturellement ce rôle de moteur dans l'animation de la suppléance, à la suite de l'action de la Diren.

Le rapport 025. SddeC a été écrit pour permettre de prendre plus substantiellement la mesure des enjeux et du travail que représente l'établissement d'un dispositif comme le schéma directeur des données de l'environnement de Corse. Il développe et articule entre eux les principes auxquels on ne peut déroger pour publier des données et propose des éléments pour accompagner l'évolution de l'administration des données vers une configuration qui réponde à ces principes. Enfin, dans son contenu même, il propose un point d'entrée pour permettre à un honnête homme utilisateur, intéressé par les questions d'environnement mais pas au fait des arcanes des technologies modernes, de se former une vision, un cadre mental qui lui permette de naviguer à la recherche d'informations. Des comptes rendus circonstanciés des discussions qui président à l'établissement de consensus sur la signification des données, sont aussi une irremplaçable information pour permettre à un nouvel arrivant, par exemple un utilisateur qui découvre le système, de comprendre sans ambiguïté le vocabulaire qui est employé, les données qu'il va manipuler.

## **-5 Conclusion**

Le schéma a adopté les protocoles d'échange et les principes d'organisation incontournables, et les a effectivement mis en œuvre et amenés à un point de maturité avancée. Techniquement, les éléments fondamentaux sont en place, et c'est déjà un accomplissement dans le paysage contemporain. La Diren a été motrice des processus.

Il incombe à la Dreal de poursuivre et d'entretenir le travail de fond du §-4, en continuant à entraîner ses partenaires vers un ensemble opérationnel de « services de données » répondant aux critères énoncés plus haut.

La clef du succès de l'entreprise réside dans le travail intime sur les données. Le caractère pionnier de l'entreprise implique un investissement important et donne une responsabilité toute particulière à celui qui aura eu le mérite de tracer le premier sillon. De plus, la multiplicité des thèmes, la spécificité du territoire où cet assemblage a lieu impliquent un travail original de spécialisation, toujours articulé avec les autres forces normalisatrices emboîtées, nationales, européennes, internationales. Le travail déjà effectué et l'effet d'entraînement qu'il produit le placent sous les meilleurs auspices.

vincent douzal

Philippe Lemoisson